**Optical** **Flow** **Analysisfor** **Pedestrian** **Counting**

A Report Submitted inPartialFulfillmentoftheRequirements for the Degree of **Bachelor** **of** **Technology** in

**Information** **Technology**

by

**Swatantra Kumar Goswami**

**(171500354)**

under the guidance of **Mr.** **Pankaj Sharma**

to the

**INFORAMATION** **TECHNOLOGY** **DEPARTMENT**

GLA University Mathura, UTTAR PRADESH (INDIA) **May,** **2019**

**UNDERTAKING**

I declare that the work presented in this report titled ―*Optical* *Flow* *Analysis* *for* *Pedestrian* *Counting*‖, submitted to the Information Technology Department, Dr. Ambedkar Institute of Technology for Handicapped, Kanpur Uttar Pradesh (India) for the award of the ***Bachelor*** ***of*** ***Technology*** degree in ***Information*** ***Technology***, is my original work. I have not plagiarized or submitted the same workfortheawardofanyotherdegree. Incasethisundertaking is found incorrect,I accept that mydegree maybe unconditionally withdrawn.

May, 2019 Mathura U.P. (India)

(Swatantra Kumar Goswami,171500354)

ii

**CERTIFICATE**

Certified that the work contained in the report titled ―*Optical* *Flow* *Analysis* *for* *Pedestrian* *Counting*‖, by *Swatanytra Kumar Goswami (171500354) ,*has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Mr. Pankaj Sharma

Information TechnologyDept.

May, 2019 Uttar Pradesh (INDIA)

iii

**Preface**

Automated methods are commonly used to count motorized vehicles, but arenot frequently used to count pedestrian. This is because the automated technologies availabletocountpedestrianarenotverydeveloped,andtheireffectivenesshasnot beenwidelyresearched.Moreover,mostautomatedmethodsareusedprimarilyfor thepurposeofdetecting,ratherthancounting,pedestrian(Dharmarajuetal.,2001; Noyce and Dharmaraju, 2002; Noyce et al., 2006). Automated pedestrian count-ing technologies are attractive because they have the potential to reduce the labor costs associated with manual methods, and to record pedestrian activity for long periodsoftimethatarecurrentlydifficulttocapturethroughtraditional methods. Data input and storage may also be less time consuming than with manual meth-ods.Muchoftheresearchonautomatedpedestriantrackingdeviceshasfocusedon pedestriandetection,notpedestriancounting. Extensivereviewsofpedestriande-tectiontechnologieswereconductedbyNoyceandDharmaraju(2002)andbyChan et al. (2006). Technologies include piezoelectric sensors, acoustic, active and passive infrared,ultrasonicsensors,microwaveradar,laserscanners,videoimaging(com-putervision),etc. Thisreport describesoneof thewaytocount movingpedestrian in thevideo.

iv

**Acknowledgements**

Wewouldliketoacknowledgeallthosewhohavehelpedusinputtingourideasand assigned work into something concrete. We are specially grateful to Mr. Kunj Bihari for assigning us this project and also for his continuous support and prompt help. Without his valuable insights and expertise, our project would not have seen the light of day.

v

**Contents**

**Preface** **iv**

**Acknowledgements** **v**

**1** **Introduction** **1** 1.1 Motivation ................................................................................................ 1 1.2 Abstract..................................................................................................... 2 1.3 Definition oftheconcepts......................................................................... 3 1.3.1 Optical Flow.................................................................................. 3

1.3.2 Erosion 4 1.3.3 Dilating 5

1.3.4 HSV Color Model.......................................................................... 6 1.3.5 ImageThresholding ...................................................................... 6 1.3.6 Image Filtering.............................................................................. 7 1.3.7 Contours7

**2** **Related** **Work** **8** 2.1 MotionFlowSegmentation....................................................................... 8 2.2 Crowd Counting.......................................................................................10 2.3 Pedestriandetectionbasedondeeplearningmodel ............................... 10 2.4 PedestriandetectionwithConvolutionalNeuralNetworks.................... 12 2.5 Stereo basedpedestriandetection .......................................................... 13 2.6 Pedestrian detection and tracking using three-dimensional LADAR

Data......................................................................................................... 15

vi

2.7 On-Boarddetectionofpedestrianintentions...........................................16

**3** **Proposed** **Work** **18** 3.1 Optical flow computation ......................................................................20 3.2 Image thresholding ................................................................................20 3.3 Morphological processing and noise removal.........................................21 3.4 Contours detection and extraction........................................................ 21

**4** **Results** **Analysis** **22**

**5** **Conclusion** **38**

**6** **Future** **Works** **39**

**References** **42**

vii

**Chapter** **1**

**Introduction**

Pedestrian counting is important in various domains, from marketing to video surveillance.Thelargenumberofpedestriansoodingthroughtheentrancepathsof transportationstations,buildingsandshoppingmallsrepresentsanongoingchal-lenge for decision makers in charge of ensuring both public safety and business improvement.Forinstance,aparticularnumberofpersonsinagivencontextmight reect an unusual and potentially dangerous situation.

**1.1** **Motivation**

Traditionally, pedestrian counting is conducted manually, which entails high cost andmayintroducehumanerror. Therefore,automaticpedestriancountinghasre-ceived much attention. Numerous approaches to automatically counting moving objects have been reported in the literature. The most widely deployed methods utilizelasersensorsandinfraredsensors,bothofwhichareinexpensiveandrobust against changes in the illumination condition due to fluctuations of lighting and weather. However,thesemethodssometimesfailtocountpedestriancorrectly,es-peciallywhentheheightsofco-existingpedestrianaresimilarorwhentheheights donotfallwithinthepresumedrange,becausesuchmethodsdependonthediffer-ence in propagation delays of reflected laser pulses or infrared light. On the other

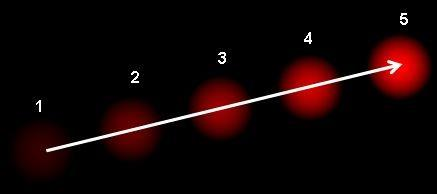
1

hand, pedestrian counting methods using video processing technology have some ad-vantagesincludingthat thecamera position is flexible, the video sequences can be retrievedremotely,andinformationsuchastheheightandgenderofpedestriancan beobtained,inadditiontothenumberofpedestrian. Thecountingofpedestrianby videoprocessingisbeingactivelyresearched,andacontestforcountingpedestrian byusingvideo-processingtechnologyisheldat theannualIEEEworkshopPerfor-manceEvaluationofTrackingandSurveillance(PETS).Texture-basedmethodsare robust to the occlusion of moving objects whose texture pattern is uniform. How-ever, pedestrian generally have non uniformity in their appearance. Model based methods use a shape model to extract moving objects, and havethe high counting accuracyinthepresenceofocclusions.However,theperformanceofthesemethods becomesconsiderablyworsewhenthemovingobjectisanon-rigidobjectsuchasa person. Insomecases,acameraissetdirectlyabovepedestriantoavoidocclusion. However, the camera position is restricted to locations such as the entrance of a building. Other methods use multiple cameras which give the three-dimensional positionsof pedestrian to identifytheir locations more accurately. However,com-puting three-dimensional positions from point correspondences over multiple images leads to high computational complexity.

**1.2** **Abstract**

Inthisreport,weproposeamethodforcountingthenumberofpedestrianinvideo sequence retrieved bya single stationary camera. Weassume a situation such asa crowdedstreetwherethedensityofpedestrianinvideoframesissufficientlyhigh. Theproposedmethodisbasedoncomputingdenseopticalflowsofmovingobjects. Wecalculateopticalflowforallthepointsinthevideoframe. Theopticalflowgives us the velocity vector of the points in the x and y direction. Obviously, the station-arypointswillhavezero(orveryless)velocity. Wethenconvertourimagetobinary image byapplying the concept of threshold on computed velocity of all the points. After binary image is obtained, we apply contours on the formed binary image to

2



count thenumberofpedestrian. Theeffectivenessof theproposedmethodiseval-uatedthroughimplementationexperimentswithseveralvideosequences. Wealso compare the performance of the proposed methodwith a background subtraction method.

**1.3** **Definition** **of** **the** **concepts**

**1.3.1** **Optical** **Flow**

Optical flow is the pattern of apparent motion of image objects between twocon-secutive frames caused by the movement of object or camera. It is 2D vectorfield where each vector is a displacement vector showing the movement of points from first frame to second. Consider the image below :-

It shows a ball moving in 5 consecutive frames. The arrow shows its displacement vector. Optical flow has many applications in areas like :

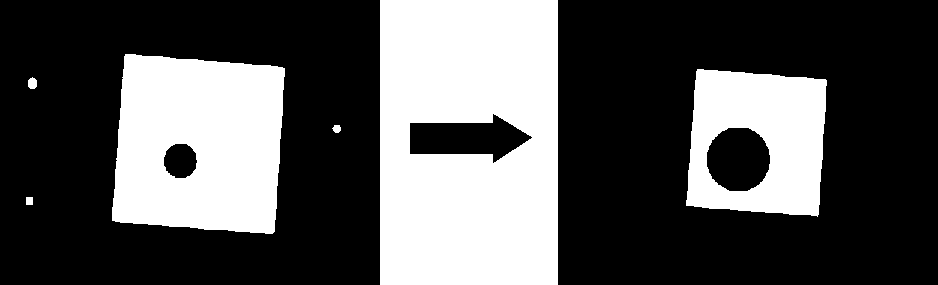
1. Structure fromMotion

2. Video Compression

3. Video Stabilization

Thepixelintensitiesofanobjectdonotchangebetweenconsecutiveframes.Neigh-bouringpixelshavesimilarmotion. ConsiderapixelI(x,y,t)infirstframe. Itmoves

3



by distance (dx,dy) in next frame taken after dt time. So since those pixels are the same and intensity does not change, we can say,

*I*(*x,* *y,* *t*) = *I*(*x* + *dx,* *y* + *dy,* *t* + *dt*)

Then take Taylor series approximation of right-hand side, remove common terms and divide by *dt* to get the following equation:

*fxu* + *fyv* + *ft* = 0

where,

*∂f* ;*fy* = *∂f* *∂x* *∂y*

*fx* =

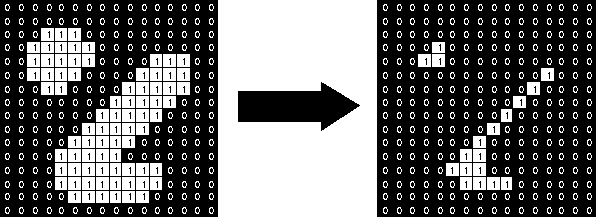
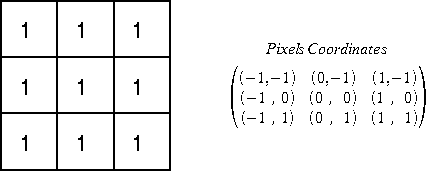
*u* = *dx*; *v* = *dy* *dt* *dt*

AboveequationsarecalledOpticalFlowequations. Theseequationsaresolved by Lucas-Kanademethod.

**1.3.2** **Erosion**

Erosionisabasicoperationanditsprimaryfeatureistoerodeawaytheboundaries of the different foreground regions. Thus this foreground objects will become smaller (littleofthemwilltotallybevanished)andholesinobjectswillbebigger. Inoutline, allthepixelsoftheforegroundobjectwhichcantotallycontainthestructureelement Bwillbecontainedintheerodedobject. Forexample,takeconsiderofa3x3square structure element having its morphological centre the same as the geometrical centre. It is as follows:

4



Tocompute a binary erosion, all the Pixels of the foreground must be process. For each pixel of the foreground, the algorithm puts the structure element (the centre of the structure element matches with the pixel) and tests if the structure elementiscompletelycontainedintheforeground. Ifitisnot,thecurrentpixelwill beconsideredlikethebackgroundandonthecontrary,ifitis,thecurrentpixelwill be contained in the eroded foreground.

**1.3.3** **Dilating**

Like the erosion, the dilatation is the second basic operation and its primary feature is to dilate the boundaries of the different foreground regions. Thus this foreground objects will become bigger and holes in objects will be smaller (little of them will totally disappear). In outline, all the pixels of the background which can touch the foreground regions, byputting on it the structure element B, will be containedinthedilatedobject. Forexample,takeconsiderofa3x3squarestructure element having its morphological centre the same as the geometrical centre (see

5

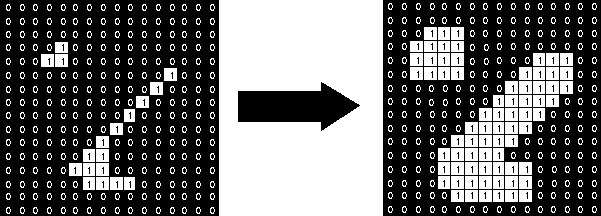


illustration 7). To compute a binary dilatation, all the Pixels of the background mustbeprocess. Foreachpixelofthebackground,thealgorithmputsthestructure element(thecentreofthestructureelementmatcheswiththepixel)andtestsifthe structure element is in touch with at least one pixel of the foreground. If it is, the currentpixelwillbeconsideredliketheforegroundandonthecontrary,ifitisnot, the pixel will stay a background pixel.

**1.3.4** **HSV** **Color** **Model**

HSVisnamedassuchforthreevalues: hue,saturation,andvalue. Thiscolorspace describescolors(hueortint)intermsoftheirshade(saturationoramountofgray) and their brightness value.

*•* Hueisthecolorportionofthecolormodel,andisexpressedasanumberfrom 0 to 360 degrees.

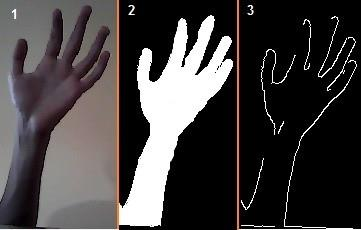
*•* Saturation is the amount of gray in the color, from 0 to 100 percent. A faded effectcanbehadfromreducingthesaturationtowardzerotointroducemore gray.

*•* Value works in conjunction with saturation and describes the brightness or intensityofthecolor,from0-100percent,where0iscompletelyblackand100 is the brightest and reveals the most color.

**1.3.5** **Image** **Thresholding**

Segmentationinvolvesseparating animage intoregionscorresponding toobjects. Weusuallytrytosegmentregionsbyidentifyingcommonproperties. Thesimplest

6



property that pixels in a region can share is intensity. So, a natural way to seg-mentsuchregionsisthroughthresholding,theseparationoflightanddarkregions. Thresholdingcreatesbinaryimagesfromgrey-levelonesbyturningallpixelsbelow some threshold to zero and all pixels above that threshold to one.

**1.3.6** **Image** **Filtering**

Image filtering is used to remove noises, sharpen contrast, highlight contours, detect edges,etc.Imagefilterscanbeclassifiedaslinearornonlinear.Linearfiltersarealso knowasconvolutionfiltersastheycanberepresentedusingamatrixmultiplication. Thresholding and image equalization are examples of nonlinear operations, as is the median filter. In our method, we have used median filtering. Median filtering is widely used as it is very effective at removing noise while preserving edges. It is particularly effective at removing ‗salt and pepper‘ type noise. The median filter works by moving through the image pixel by pixel, replacing each value with the medianvalueof neighbouring pixels. The patternof neighboursiscalledthe‖win-dow‖, which slides, pixel by pixel, over the entire image. The median is calculated byfirstsortingallthepixelvaluesfromthewindowintonumericalorder,andthen replacing the pixel being considered with the middle (median) pixel value.

**1.3.7** **Contours**

Contourscanbeexplainedsimplyasacurvejoiningallthecontinuouspoints(along the boundary), having same color or intensity. The contours are a useful tool for shape analysis and object detection and recognition. Contour tracing is appliedto digital images in order to extract their boundary.

7

**Chapter** **2**

**Related** **Work**

Formore than 40 years researchers havebeen studying pedestrian dynamics with the aim of supporting the design of pedestrian facilities. Since the population of the world is increasing and concentrating in urban areas, and due to the growing relevance of mass events (like sports contests, concerts and festivals periodically arranged and attracting growing number of pedestrian from different parts of the world), adequate safety measures are becoming ever more important. More recently, researchersarefocusingonstudyingcrowddynamicsinordertoimprovetheevac-uation strategies in emergency situations.

**2.1** **Motion** **FlowSegmentation**

An important contribution that automated analysis tools can give to pedestrian and crowd safety is the detection of conflicting large pedestrian flows: this kind of movement pattern, in fact, may lead to dangerous situations and potential threats to pedestrian‘ safety. Therefore, segmenting typical flow patterns of crowd and estimating the number of pedestrian in crowd are important steps to understand overall crowd dynamics. Crowd flow segmentation has multiplebenefits:

1. Enables clutter free visualization of moving groups

2. It is independent from ―detection and tracking‖

8

3. Provides input for the pedestrian simulation models (in terms of data for simulation initialization or validation)

Automatic analysis of the crowd has become the center of focus for most of re-searchers in computer vision. Detecting pedestrian and tracking are traditional waysof crowdanalysis. Most algorithmsdevelopedforobject detection andtrack-ingworkwellwithpedestrianinlowdensitycrowdswherethenumberofpedestrian is generally less than twenty individuals in a single frame, but with higher densi-ties (where the number of pedestrian in a frame can be in the order of hundreds), detection and tracking of individuals are almost impossible due to multipleocclu-sions. Therefore, the research has focused on gathering global motion information at higherscale. Globalanalysisofdensegroupofmovingpedestrianisoftenbasedon optical flow analysis. [AS07] proposed particle dynamic segmentation of crowd flows bydetectingthelagrangiancoherentstructuresoverthephasespace. Buttheirpro-posedmethodiscomputationallyexpensivebecauseofthecalculationofFTLEand also could not detect small flows. [OYA10] used SIFT features to detect dominant motion flows. Flow vectors of SIFT features are calculated and then motion flow map is divided into small regions of equal size. In each region, dominant motion flowsareestimatedbyclusteringflowvectors.[EB08]proposedspectralclustering methodforcrowdflowsegmentationbycomputingsparseopticalflowfield.Crowd flow is estimated using multiple visual features reported by[SND11] where flow is estimated by the number of persons passing through a virtual trip wire and accu-mulatethetotalnumberofforegroundpixels. Min-cut/maxflowalgorithmisused by Ullah et al. [UC12] for crowd flow segmentation. In all of above four methods, wecannot findclear boundaries amongdifferent flows. Crowdflow segmentation byusinghistogramcurvesisreportedby[LRZ12]whereanglematrixofforeground pixelsissegmentedinsteadofopticalflowforeground. Thederivativecurveofhis-togram is used to segment the flow. Since this method only looks to the peaks of histogram curve, therefore it loses information about the crowd flow.

9

**2.2** **Crowd** **Counting**

Most of the literature in field of crowd density estimation has focused on segmen-tation of pedestrian or head counts. Some of the work focused on texture anal-ysis or wavelet descriptor for estimating crowd density. Bayesian based segmen-tation proposed by [ZN03] to segment and count the pedestrian but this method fails in high density scenarios as because of severe occlusions. [YST10] extract blob features of moving objects and neural network is trained to estimate num-ber of pedestrian in each blob. [XLH06] classified crowd density into four classes byusingwaveletdescriptors. [MHL08]usedtexturedescriptorstoestimatecrowd density. [TYOY99] count the number of pedestrian as they cross some virtual line. [HMY+97]usedinfra-redimagingtocountthenumberofpedestrianincrowd.Sim-ple background subtraction from static images to estimate the crowd densitywas proposedby[RP07].Backgroundremovalconceptisusedtoestimatethecrowdarea by[VYD+93, VYD+94]. [RMAS04] used a forward facing camera mounted on the car to detect crowd of pedestrian. Trained support vector machines using HAAR transform is used by[LCC01] to identify heads of pedestrian. Median background computing techniques used by [RP07] to extract foreground pixels. Support vec-tor machine, K-nearest neighbour, PNN, BPNN are used to classify images in two categories(zeropersons,oneormorepersons). Asthesensorarebecomingcheap, therefore,recentlymanyresearcherscountthepedestrianusinginfra-redsensors. [TS07] proposed lightweight camera sensor nodes to count the pedestrian in the indoorenvironmentbasedonmotionhistogram. Recentlymanyinfra-redsensors specifically designed for pedestrian counting are available in the market.

**2.3** **Pedestriandetectionbasedondeeplearning** **model**

PedestrianDetectionistodeterminewhethertheimagecontainsapedestrianand mark the specific location of pedestrian traveling. With the development of com-puter vision, pedestrian detection in the intelligent auxiliary driving, intelligent

10

monitoring, pedestrian analysis and intelligent robot, and other fields have been widelyused. However,becauseofthecomplexityofthereallifebackground,pedes-trian posture diversity and diversification of shooting Angle, pedestrian detection is an open challenge which calls for precise algorithms. Pedestrian detection al-gorithms can be roughly divided into twoclasses: based on background modeling methodandthemethodbasedonstatisticallearning. In2003,apedestrian detec-tion algorithm based on box-shaped filters was proposed by Viola and Jones. The algorithm uses AdaBoost method to classify the features from a large number of simple Haar features, and obtains good results. This is a breakthrough in the field of target detection, and the method was succeeded used in pedestrian detection. In2005,Dalal et al. proposed HistogramOf Gradients (HOG) [231] feature. HOG featureisaveryeffectivefeature,combinedwithasimplelinearSupportVectorMa-chine(SVM),andachievedverygoodresults,atthetimeofthepedestriandatabase MIT test set on the detection rate can reach 100%, to this end, Dalai established A pedestriandatabaseINR1Aforeverydaylife,whichgreatlyincreasedthedifficulty ofpedestriandetection. Thedevelopmentofpedestriandetectionhasbeengreatly promoted by the introduction of HOG feature, which has a far-reaching impact onthesubsequentresearch. Mostofthearticlesonpedestriandetectionhavebeen extendedonthisbasissincethen.Inordertoimprovethespeedofpedestriandetec-tion, Zhu et al. used the histogram technique to calculate the HOG feature quickly and the speed was nearly 70 times higher than that of literature 2. Felzenswalb et al. [2311] proposed Deformable Part Model (DPM) algorithm bycombining the HOG,whichfurtherimprovedthereallydetectionaccuracy. Themodelisaweakly supervised algorithm that does not need to provide the location category marker becauseithasbeenconsideredasanunknownparametertoparticipateintraining. DPM takes intoconsideration the target dueto theinternal structure, and can test differentstanceofpedestriansandcandistinguishbetweentargetandbackground very well. Dollar et al. [232] calculated the Hal feature on multi-channel such as gradientdirectionandgradientamplitudeafterquantization,andthenconstructed the system frame of a simple multi-feature fusion. At the same time, the integral channelcharacteristicwasproposed,whichcanquicklycalculatethemulti-feature

11

channelofimage,suchaslocalsum,histogramandHalfeature,andrealizethefast operation of multi-feature fusion.

**2.4** **Pedestrian** **detection** **with** **Convolutional** **Neu-ral** **Networks**

The structure of the pedestrian detection systems described in the literature can roughly be divided into region of interest (ROI) detection,feature extraction, can-didate classication and tracking. Region of interest detection is most frequently based on stereo vision [241][242] or hot-spot detection in far infrared images [243]. The feature extraction module is different in almost all systems, as illustrated in Table I. The common point is that the features are always manually designed, for example wavelet coefcients or edge strength coefcients. The most frequently used classier is the support vector machine(SVM) due to its high generalizationability [242] [243] [244] [245], but there are also systems using neural networks [246], time delay neural networks [247], boosted combination of linear classiers [248][249] and templatematching[2410]. Trackingis usually based on Kalman-lter[242], mean-shifttracker[243]oralpha-betatracker[2411]. Webelievethatthekey-component ofthedetectionsystemistheclassier. EventhoughtheROIDandtrackeralsoplays animportantroleinimprovingcomputationalefciencyandaccuracy,theyaremore or less independent from the classier.

Traditionally, feature extraction and classication have been treated indepen-dently; the role of the feature extractor being considered to reduce the dimensionality of the input vector before classication. Recent research in the object detection lit-erature indicates, however, that the features play a critical role for high accuracy classication [2412][2413]. [2413] has shown that automatically derived features signi-cantly improve the accuracy of a face detection system modeled after the mammalian vision system. The cascade-based detection architecture of Viola and Jones [248], as wellasthedetectionsystemofSchneiderman[2414]alsorelyonautomaticallyse-lected features for efcient, high accuracy detection. Although these systems consider

12

the problem of automatically choosing features that are best suited for thedetec-tiontask,featureoptimizationislimitedtofeatureselectionfromapredetermined discrete set, such as Haar-wavelets in [2414] or the rectangle lters in [248].

The convolutional neural network (CNN) architecture proposed by LeCun [2415], ontheotherhand,treatsthefeatureexctractorandtheclassicationcomponentiden-tically. The feature extraction lters are implemented as a hidden layer with shared weightsthatareoptimizedtogetherwiththeweightsoftheclassicationcomponent so that the total classication error is minimized. Since the optimization is based onagradient-basedminimizationprocedure,theresultingfeatureextractionlters canhavearbitrarycontinuouscoefcients,notlikethebinaryrectangularltersused in [248][2414]. CNNs have been reported to achieve high accuracy at a low com-putational cost in several image recognition problems, including character recogni-tion [2415][2416], hand-tracking[2417], face detection [2418][2419], face recognition [2420], facial expression recognition [2421] and generic object classication [2422]. Although the CNN has several favorable characteristics, it has been employed by still relatively few research groups. Weconjecture that, besides the lack of a CNN implementation in most widely used neural network simulators, this is due to the small number of studies systematically comparing CNN-s with other classiers. It is, therefore, one purpose of this paper to provide a well controlled comparison of CNN with SVM, the most popular classier in recent pedestrian detection systems. We demonstrate that CNN-sprovide superior detection accuracy to SVM-s using classicimagefeatures. AlthoughCNN-shavebeenpreviouslyappliedwithsuccess to frontal face detection, thisis the rst demonstration of applicability toobject de-tection that requires contour information in an unconstrained natural environment.

**2.5** **Stereo** **based** **pedestrian** **detection**

A stereo-based generic object detection algorithm to generate initial pedestrian can-didatewindowsexclusivelyfromrangemaps.Thegroundplane(XZinourconven-tion)isdividedintoadiscreteregular2Dgrid.Eachspecificgridlocationrepresents a3Dgroundplanelocationw.r.t. thecamera,andapedestrianatthis positioncan

13

bedescribedbyasingle3Dgeometricdescriptionofitsshape.Apedestrian-shaped cuboid rendered at the given location gives us a depth-template representation of thepedestrianasseenfromthecameraviewpoint. Thedepthtemplatesforallgrid locationsarepre-computedforefficientmatchingatrun-time.Atrun-time,forany given frame, we have the three disparity maps Di, i = 0, ..., 2 available. For each 3D template, we look at its distance range Z and then select one of the levels Di to perform 3D template matching. The level is selected to ensure that, at each loca-tion, only the relevant resolution disparity map with sufficient details is used. We use D2 from 0 5 m, D1 from 5 15 m and D0 beyond 15 m. For the matching step, the template is correlated with the appropriate level Di bysearching around the X and Z directions, and around the Y (vertical) direction to account for local pitch uncertainty due to calibration errors and bumps in the road surface. The output of this template matching is a correlation score map (over the horizontal 2D grid) from which peaks are selected by non-maximal suppression as in [2517]. Around eachpeak,theareaofthecorrelationscoremapwithvalueswithin60%ofthepeak score is projected into the image to get the initial pedestrian ROI candidate set. Notethat thisdetection stagemust ensureverysmall pedestrianmiss rates,hence a larger number of peaks obtained bynon-maximal suppression is acceptable. We relyon additional steps detailed next toprunethese candidates. The initial pedes-trianROIcandidatesetispruned,first,byconsideringtheoverlapbetweenmultiple ROIs: detection with more than 70% overlap with existing detection are removed. After this pruning step, a Canny edge map is computed for each initial pedestrian ROI.Theedgepixelsarefilteredusingthedisparitymaptoremovepotentialback-groundedgesbyconsideringadisparityrangearoundthedisparityofthedetected ROI.Averticalprojectionofthebinarymaskcorrespondingtotheremainingedges results in a 1D histogram from which peaks are detected using mean-shift. Each such peak potentially corresponds to a pedestrian due to the high density of edges on pedestrian. A new pedestrian ROI is initialized at each detected peak, which is refined first horizontally, and then vertically to get a more centered and tightly fittingboundingboxonthepedestrian. Therefinementprocessinvolvesusingver-ticalandhorizontalprojections,respectively,ofbinarizeddisparitymaps(similarto

14

using the edge pixels above) followed by detection of peak and valley locations in the computed projections. After these refinements, any resulting overlapping detec-tions are again removed from the detection list. Withthis approach, wecan detect pedestrian upto a range of 40 m.

**2.6** **Pedestrian** **detection** **and** **tracking** **usingthree-dimensional** **LADAR** **Data**

The approach investigated in this work employs three-dimensional LADARmea-surements to detect and track pedestrian over time. The sensor is employed on a movingvehicle. Thealgorithmquicklydetectstheobjectswhichhavethepotential ofbeinghumansusingasubsetofthesepoints,andthenclassifieseachobjectusing statistical pattern recognition techniques. The algorithm uses geometric andmo-tionfeaturestorecognizehumansignatures. Theperceptualcapabilitiesdescribed formthebasisforsafeandrobustnavigationinautonomousvehicles, necessaryto safeguard pedestrian operating in the vicinity of a moving robotic vehicle.

The ability to avoid colliding with other objects is essential in autonomous ve-hicles,especiallyin caseswheretheyoperate incloseproximityto pedestrian. The timely detection of a pedestrian makes the vehicle aware of a potential danger in its vicinity, and allows it to modify its course accordingly. There is a large bodyof workdoneusinglaserlinescannersastheprimarysensorforpedestriandetection andtracking.Inourgroup,wehavedevelopeddetectionandtrackingsystemsusing SICKTM laserlinescanners;theseimplementationsworkwell insituationswhere the ground is relatively flat [265]. However, a 3D LADAR (i.e. one who produces a set of 3D points, or point cloud) captures a more complete representation of the environment and the objects within it. An algorithm was presented that detects pedestrianfrom3Ddata. Itsmainimprovementovertheversionwith2Ddatawas that it constructsa ground elevation map,anduses it toeliminate groundreturns. Thisallowspedestrian detectionevenwhenthesurroundinggroundisuneven. To classify the humans the algorithm uses motion, size, and noise features. Persons are classified well as long as they are moving. However, there are still too many

15

falsepositiveswhenclassifyingstationaryhumans. Astrategytodetectandclassify humansusingthefull3Dpointcloudoftheobject. Thiswillimprovetheclassifica-tionof bothmovingandstaticpedestrian.However,theimprovement willbe most significant for static humans. The algorithm quickly detects the objects that have thepotential ofbeinghumansusingasubset ofthepointcloud,andthenclassifies eachobjectusingstatisticalpatternrecognitiontechniques.In[268]theauthorsre-port an algorithm capable of detectingboth stationary and moving humans. Their approach uses multi-sensor modalities including 3D LADAR and long wave infrared video(LWIR).Similarly,in[269]thesameresearchgrouppresentsatechniquefor detecting humans that combines the use of 3D LADAR and visible spectrum im-agery. Inbotheffortstheauthorsemploya2Dtemplatetoextractfeaturesfromthe shape of an object. Among other differences, as opposed to our work, they extract ashapetemplatefrom theprojection inonlyoneplane,andcompute ameasureof howuniformlydistributedthereturnsareacrossthetemplate.PedestrianDetection and Tracking Using Three-Dimensional LADAR Data.

**2.7** **On-Board** **detection** **of** **pedestrianintentions**

Avoiding vehicle-to-pedestrian crashes is a critical requirement for nowadays ad-vanced driver assistant systems (ADAS) and future self-driving vehicles. Accord-ingly,detectingpedestrianfromrawsensordatahasahistoryofmorethan15years of research, with vision playing a central role. During the last years, deep learning hasboostedtheaccuracyofimage-basedpedestriandetectors. However,detection isjustthefirststeptowardsansweringthecorequestion,namelyisthevehiclegoing to crash with a pedestrian provided preventive actions are not taken? Therefore, knowingassoonaspossibleifadetectedpedestrianhastheintentionofcrossingthe road ahead of the vehicle is essential for performing safe and comfortable maneuvers thatpreventacrash.However,comparedtopedestriandetection,thereisrelatively little literature on detecting pedestrian intentions. This paper aims to contribute alongthislinebypresentinganewvision-basedapproachwhichanalyzesthepose of a pedestrian along several frames to determine if he or she is going to enter the

16

roadornot.Wepresentexperimentsshowing750msofanticipationforpedestrian crossing the road, which at a typical urban driving speed of 50 km/h can provide 15additionalmeters(comparedtoapurepedestriandetector)forvehicleautomatic reactionsortowarnthedriver.Moreover,incontrastwithstate-of-the-artmethods, ourapproachismonocular,neitherrequiringstereonoropticalflowinformation.

17

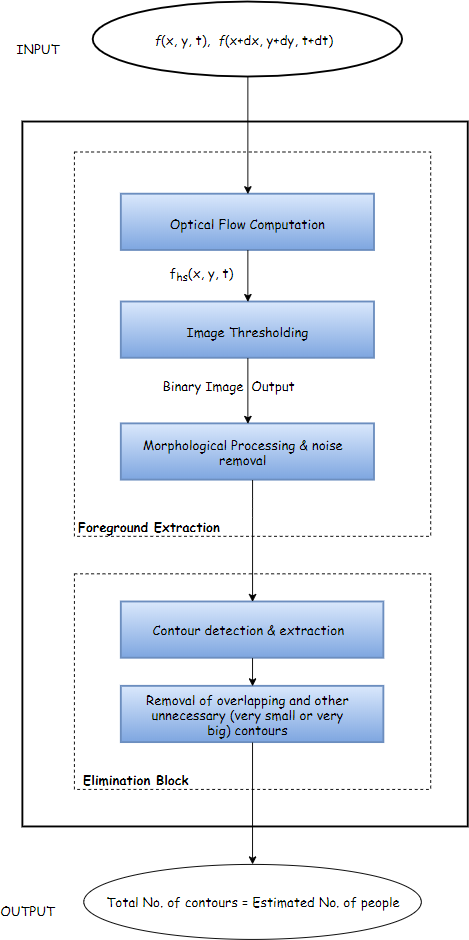
**Chapter** **3**

**Proposed** **Work**

Usuallycrowdedobjectsmoveinwideareas,andtodetectchangeinpixels,optical flowmethodscanbeused. Therearemainlytwoopticalflowalgorithms: (a)Lucas-Kanade optical flow, and (b)Dense optical flow.

Since, there is need to detect change in every pixel, hence dense optical flow method is used. Since the optical flow vector of each pixel has the magnitude and direction values, we use magnitude information to extract the foreground, all the pixelswhichhavehighermagnitudethanthresholdwillbeclassifiedasforeground. Consider the following flow chart :

18



19

Our proposed framework is composed of four main processing blocks:

1. Optical flowcomputation

2. Image thresholding

3. Morphological processing and noise removal

4. Contours detection and extraction

Inthissection,eachprocessingblockisdiscussedindetail. Forsakeofdescription oftheproposedapproach,videostakenfromacrowdrelateddatasetareemployed.

**3.1** **Optical** **flowcomputation**

Computationofopticalflowisthemostimportantpreprocessingstepfordetecting themovingobjectsfromthevideoandthereforeformsthebasisoftheframework. Optical flow computation is useful for detection, tracking and understanding the behavior of the object. Traditionally, in video surveillance with a fixed camera, researchersusebackgroundsubtractionmethod,whereforegroundobjectsareex-tracted from video if the pixels in the current frame deviate significantly from the background. In this report, foreground mask, that is generated by optical flow, *fhs*(*x,y,t*) is used. Two consecutive frames *f*(*x,y,t*) and *f*(*x*+ *dx,y*+*dy,t*+*dt*) areappliedtoforegroundextractionblock.Thenthedenseopticalflowiscomputed betweentheadjacentframes.Foreachframe,aftercalculatingopticalflow,velocity vectors of all the points in the 2-D plane are obtained.

**3.2** **Image** **thresholding**

Fromtheabovestepvelocityvectorsofallthepointsinthe2-Dplaneareobtained. Then, the velocity of each point in the frame is computed. Now the image is con-vertedtothehsvimage,insuchawaysuchthatthevalueofv(whichistheintensity ofhsvimage)isdirectlycorrelatedwiththevelocityofthepoints. Thepointswhich arestationarywillhavezero(orveryless)velocityandhencetheywillhaveveryless

2 0

(or zero) intensity in the hsv image and vice versa. The obtained hsv image in the abovestepisnowconvertedintothebinaryimagebyusingtheconceptofthreshold. Thevalueof thresholdwhichis takenis zero. Thus, thepointswhich will havezero intensitywill be assigneda valueof 0(black) andrest of thepointswill be assigned value1(white). Since,fromtheaboveobtainedhsvimageonlystationarypointswill havezerointensity,henceafterapplyingimagethresholding,stationarypointswill be given a 0 value(black) and moving points will be given a 1 value(white). After theabovestepwhitepatchesintheblackbackgroundwillbeobtained. Thosewhite patches denotes moving objects.

**3.3** **Morphologicalprocessingandnoiseremoval**

After the binary image is obtained, still there will be a lot of noise in the image andthustoremovethenoiseweapplymorphologicalprocessingfollowedbymedian filter.

**3.4** **Contours** **detection** **and** **extraction**

Now, the contours are created on the binary image obtained from the above step. Butsince,noisewasnotremovedfullyfromthebinaryimage,hencealotofunneces-sarycontourshavingaverysmallareawillalsobeformed. Also,lotsofoverlapping contours will also be formed. This step focuses on removing such contours and otherunnecessarycontours(verybigonesorverysmallones). Finally,thenumber ofcontoursinthecurrentframe,willgiveustheapproximatenumberofpedestrian in the current frame.

21

**Chapter** **4**

**Results** **Analysis**

This section presents the quantitative analysis of the results obtained from exper-iments. The experiments were carried out on a PC of 2.7 GHz (Core 7) with 8.0 GB memory. The data set coverstwotypes of crowded scenarios: the first scenario consists of videos involving high density crowds. The second scenario covers low density crowds i.e. road crossing video, where pedestrian are moving over zebra crossingin different directions,androad video,wherevehiclesandpedestrian are moving in different directions on road.

22



1. Video 1 : One Cross way traffic (High density)

(a) Background Subtraction

Figure 1: Original frame

Figure 2: After Background Subtraction

23



Figure 3: Final Image

24



(b) Optical Flow

Figure 4: Original frame

Figure 5: Optical flow frame

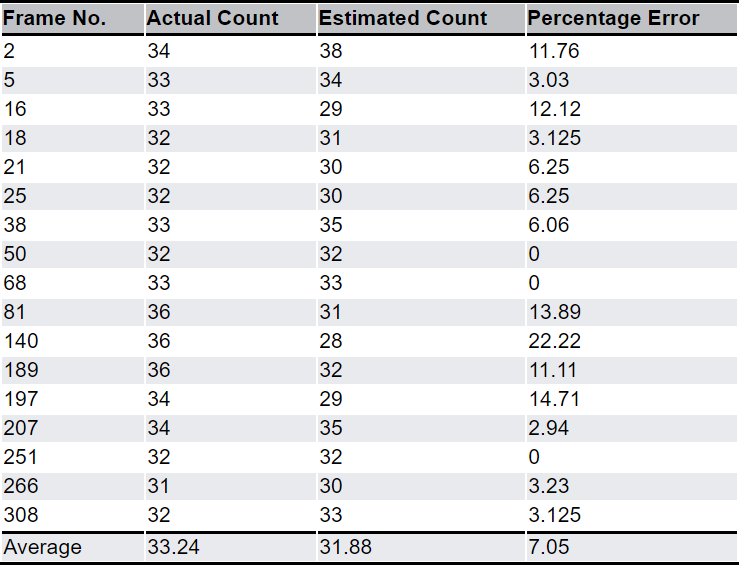
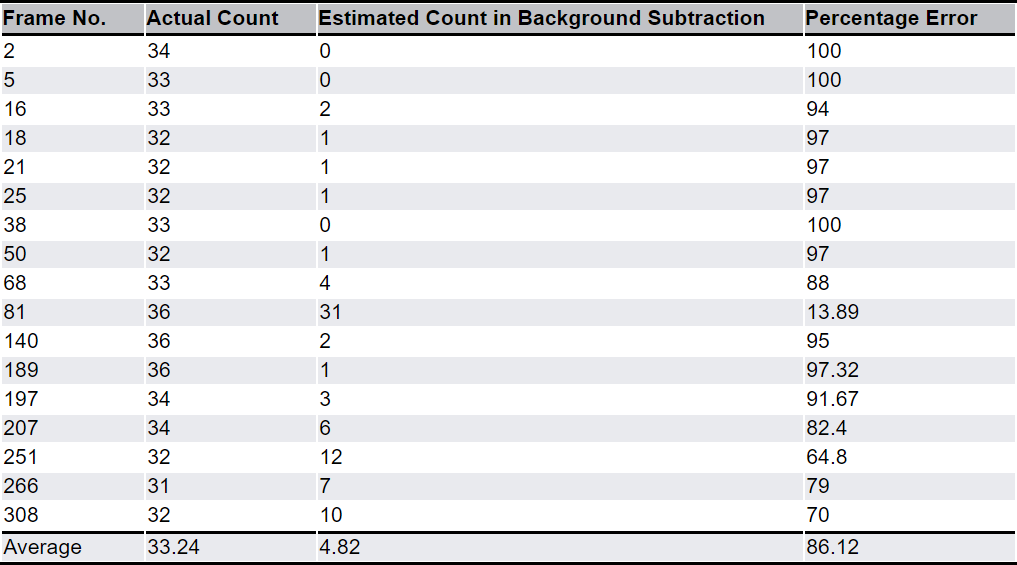
25



Figure 6: Binary Image

Figure 7: Resultant Contour

26



(c) Comparison

Figure 8: Result table of Background Subtraction

Figure 9: Result table of optical flow

27



2. Video 2 :Four ways traffic (High density)

(a) Background Subtraction

Figure 10: Original frame

Figure 11: After Background Subtraction

2 8



Figure 12: Final Image

29



(b) Optical Flow

Figure 13: Original frame

Figure 14: Optical flow frame

3 0

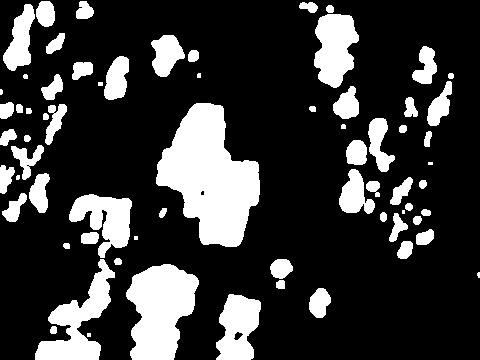
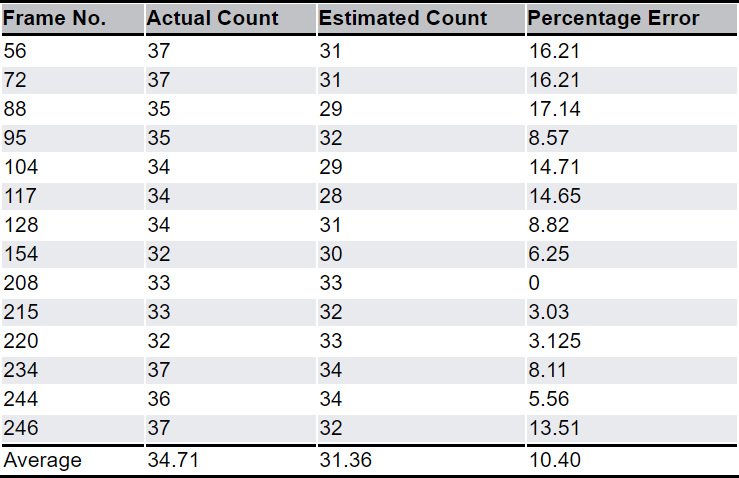
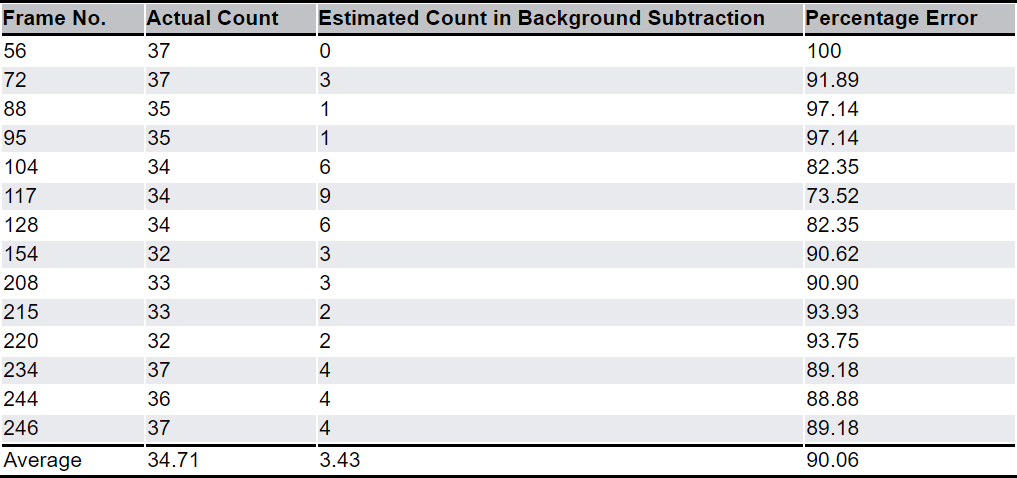


Figure 15: Binary Image

Figure 16: Resultant Contour

31



(c) Comparison

Figure 17: Result table of Background Subtraction

Figure 18: Result table of optical flow

32



3. Video 3 :Office Video (Low density)

(a) Background Subtraction

Figure 19: Original frame

Figure 20: After Background Subtraction

33

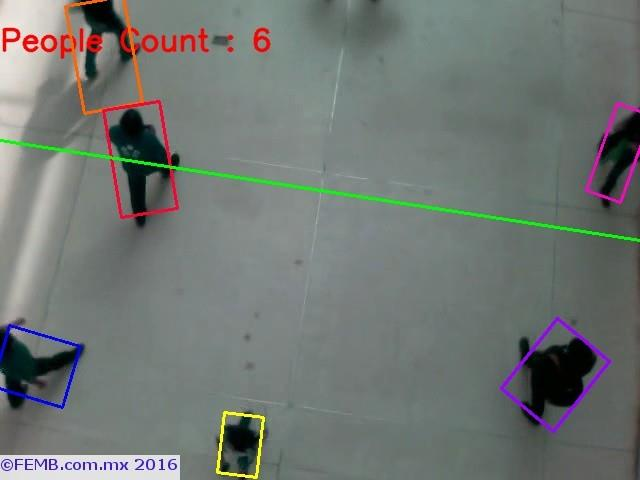


Figure 21: Final Image

34



(b) Optical Flow

Figure 22: Original frame

Figure 23: Optical flow frame

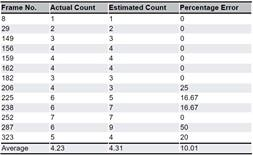
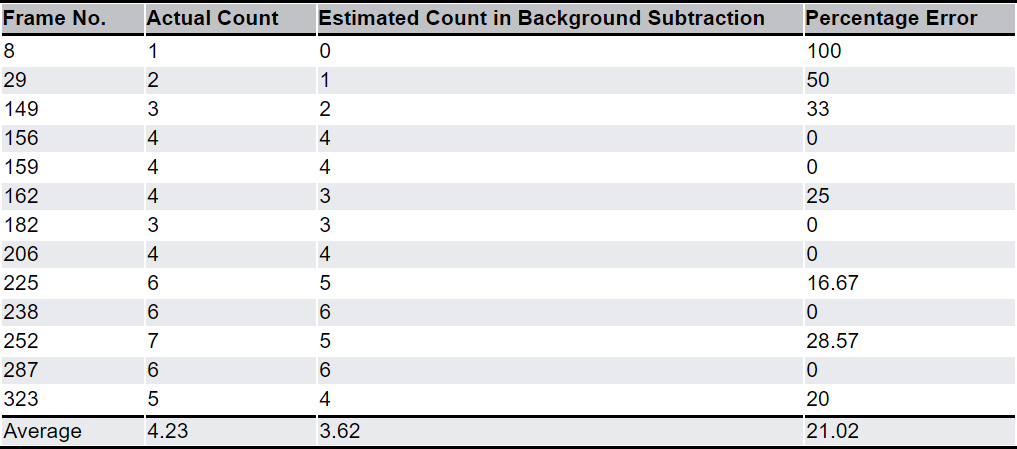
35



Figure 24: Binary Image

Figure 25: Resultant Contour

36



(c) Comparison

Figure 26: Result table of Background Subtraction

Figure 27: Result table of optical flow

37

**Chapter** **5**

**Conclusion**

It can be clearly seen from the above obtained comparison tables that optical flowmethodprovidedagreateraccuracythanthebackgroundsubtractionmethod. Also,inthispaper,bothhighandlowdensitycrowdsareconsideredandaframework that automatically counts the number of pedestrian in each frame was proposed. Such kind of analysis provides a useful input to pedestrian simulation models. A firstemploymentofanalysisoftheprojectisrelatedtotheactualinitialconfigura-tion of the simulation scenario. Second way to exploit data resulting fromproject analysisisrepresentedbypedestriancountinganddensityestimation(sincethein-dication of the average number of pedestrian present in the simulated portion of the environment is important in configuring the start areas). Finally, the above analysis can be used in the validation of the simulation results. The approach pre-sented in this report is applicable in many different situations. This method does notrequirerecognitionandtrackingofpedestrian,hencepreservingtheprivacyof the pedestrian.

3 8

**Chapter** **6**

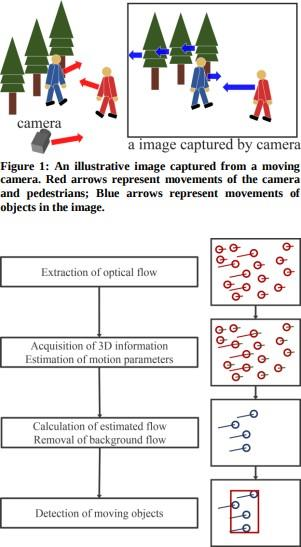
**Future** **Works**

This project wasaimed to count the number of pedestrian in the video frame. But, due to complex background information, shadow and occlusions, it is difficult to counttotalpersonsaccurately. Manydifferentadaptations,tests,andexperiments havebeen left forthe future due tolack of time(i.e. the experiments with real data are usually very time consuming, requiring more time). Future work concerns deeper analysis of particular mechanisms, new proposals to try different methods, or simply curiosity. Forthis project to work properly inthe real world further improvements can be made. This work can be extended to also count the persons who are not moving. Also, accuracy of the predictions in case of high occlusions and shadows can be improved. The scenarios, techniques and concepts which can be appliedin order to extend this project are :

1. In order to get both moving pedestrian and stationary pedestrian, a block-updating way to update the background model can be designed. Secondly,a multi-viewhead-shouldermodelcanbetrainedtofindcandidatepedestrian, and an improved k-means clustering can be used to locate the position of each pedestrian. Finally, a temporal filter with frame-difference can be used torefinethecountingresultsanddetectnoise,suchasdouble-count,random disturbance.

2. The scenario where camera is also moving can also be considered. First, the cameramotionparameterscanbeestimatedbyusingopticalflowwithastereo

39



camera.Second,theopticalflowoccurringinthebackgroundcanberemoved. Finally,movingobjectscanbedetectedindividuallybylabelingtheremaining optical flow.

3. Nowadays, cyclist accidents are also increasing in number. In comparison to other road users, such as cars and pedestrian, the automated cyclist data collection is relatively a new research area. In future work, a vision-based method for gathering cyclist count data at intersections and road segments canbedeveloped. First,amethodologyforanefficientdetectionandtracking of cyclists can be developed. A Convolutional Neural Network (CNN) based detectorcalledYouOnlyLook Once(YOLO) canbe implementedtoincrease the detection accuracy. The trajectory data obtained can be further utilized for cyclist behavioral modeling and safety analysis.

4 0



4. Also sometimes, detecting humans is not possible in dense crowds due to severe occlusions. Only the heads may be visible at this scale. So we can estimate thecountbydetectingonlyheadsintheimage. Textureanalysismethodsand SIFT-based analysis can be used for adding robustness to the system. Each detectioncanbeaccompaniedbythescaleandconfidenceassociatedwithit.

5. VirtualGate: Thisprojectcanbeexpandedtoavision-basedreal-timepedes-trian counting system for complex situations that is not dependent on an existingbuildingbackgroundmodel.Furthermore,itcanbehopedthatusers canspecifyanymovingdirectionofpedestrianpassingthroughavirtualgate. Ifapersonpassesthroughthegateinaspecificdirectionthatusersappointed in advance, this system will capture image immediately and analyze it for other applications.

41

**References**

[AS07] Saad Ali and Mubarak Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *CVPR*. IEEE Computer So-ciety, 2007.

[OYA10] OvguOzturk,ToshihikoYamasaki,andKiyoharuAizawa.Detectingdom-inant motion flows in unstructured/structured crowd scenes. In ICPR, pages 3533—3536. IEEE,2010.

[EB08] Gu¨nther Eibl and N Brandle. Evaluation of clustering methods for finding dominant optical flow fields in crowded scenes. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, pages 1—4. IEEE, 2008.

[SND11] SatyamSrivastava,KaKiNg, andEdwardJ.Delp.Crowd flow estimation usingmultiplevisualfeaturesforsceneswithchangingcrowddensities.InAVSS, pages 60–65. IEEE Computer Society, 2011.

[UC12] Habib Ullah and Nicola Conci. Crowd motion segmentation andanomaly detectionvia multi-label optimization. In ICPR workshopon PatternRecogni-tion and Crowd Analysis, 2012.

[LRZ12] WeiLi,Jiu-HongRuan,andHua—AnZhao.Crowdmovementsegmenta-tionusingvelocityfieldhistogramcurve.InWaveletAnalysisandPatternRecog-nition (ICWAPR), 2012 International Conference on, pages 191–195. IEEE, 2012.

[HMY+97] Kazuhiko Hashimoto, Katsuya Morinaka, Nobuyuki Yoshiike, Chji-hiro Kawaguchi, and Satoshi Matsueda. pedestrian count system using

42

multi—sensingapplication.InSolidStateSensorsandActuators,1997.TRANS-DUCERS‘97 Chicago., 1997 International Conference on, volume 2, pages 1291–1294. IEEE,1997.

[LCC01] Sheng-Fuu Lin, Jaw—YehChen, and HungXin Chao. Estimation of num-berofpedestrianincrowdedscenesusingperspectivetransformation.Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 31(6):645—654, 2001.

[LWMZ10] Wei Li, Xiaojuan Wu, K Matsumoto, and Hua-An Zhao. Crowd fore-grounddetectionanddensityestimationbasedonmoment.InWaveletAnalysis and Pattern Recognition (ICWAPR), 2010 International Conference on, pages 130—135. IEEE,2010.

[MHL08] WenhuaMa,LeiHuang,andChangpingLiu.Advancedlocalbinarypat-tern descriptors for crowd estimation. In Computational Intelligence and Indus-trial Application 2008.PACIIA‘08.Pacific-AsiaWorkshopon,volume2,pages 958–962. IEEE, 2008.

[RMAS04] Pini Reisman, Ofer Mano, Shai Avidan, and Amnon Shashua. Crowd detection in video sequences. In Intelligent Vehicles Symposium, 2004 IEEE, pages 66—71. IEEE, 2004.

[RP07] DamianRoqueiroandValeryAPetrushin.Countingpedestrianusingvideo cameras.TheInternationalJournalofParallel,EmergentandDistributedSys-tems, 22(3):193—209, 2007.

[TS07]ThiagoTeixeiraandAndreasSavvides.Lightweightpedestriancountingand localizing in indoor spaces using camera sensor nodes. In Distributed Smart Cameras,2007.ICDSC‘07.FirstACM/IEEEInternationalConferenceon,pages 36—43. IEEE,2007.

[TYOY99] K Terada, D Yoshida, Shunichiro Oe, and J Yamaguchi. A method of counting the passing pedestrian by using the stereo images. In Image Processing,

43

1999. ICIP 99. Proceedings. 1999 International Conference on, volume 2, pages 338—342. IEEE, 1999.

[VYD+93] SA Velastin, JH Yin, AC Davies, MA Vicencio—Silva, RE Allsop, and A Penn. Analysis of crowd movements and densities in built-up environments using image processing. In Image Processing for Transport Applications, IEE Colloquium on, pages 8—1. IET, 1993.

[VYD+94] SA Velastin, JH Yin, AC Davies, MA Vicencio-Silva, RE Allsop, and A Penn. Automated measurement of crowd density and motion using image processing. In Road Traffic Monitoring and Control, 1994., Seventh International Conference on, pages 127—132. IET, 1994.

[XLH06] Li Xiaohua, Shen Lansun, and Li Huanqin. Estimation of crowd density based on wavelet and support vector machine. Transactions of the Institute of Measurement and Control, 28(3):299–308, 2006.

[YST10] Satoshi Yoshinaga, Atsushi Shimada, and Rin—ichiro Taniguchi. Real-timepedestriancountingusingblobdescriptor.ProcediaSocialandBehavioral Sciences, 2(1):143—152,2010.

[231] N. Dalal, B.Triggs,‖Histograms of oriented gradients for human detection‖, Computer Vision and Pattern Recognition 2005 .CVPR 2005. IEEE Computer Society Conference, vol. 1, pp. 886-893, June 2005.

[232] P.Doll‘ar,Z.Tu,P.Perona,S.Belongie,‖Integralchannelfeatures‖,Proceed-ings of the British Machine Vision Conference. BMV A Press, pp. 91.1—91.11, 2009.

[2311] P.Felzenszwalb,R.Girshick,D.McAllester, D.Ramanan, ‖Object detection with discriminatively trained part-based models‖, Pattern Analysis and Machine Intelligence IEEE Transactions, vol. 32, no. 9, pp. 1627—1645, Sept2010.

[241] D. M. Gavrila and S. Munder, ―Vision-based pedestrian protection: The PROTECTOR system,‖ in Proc. IEEE Intelligent Vehicle Symposium, IV2004, Parma, Italy, June 2004.

44

45